

## 修 士 論 文 の 和 文 要 旨

研究科・専攻	大学院 情報理工 学研究科 情報・ネットワーク工学 専攻 博士前期課程		
氏 名	五十嵐 棕介	学籍番号	1731015
論 文 題 目	事前分布の推定に基づく無歪情報源符号 ～符号化アルゴリズムと冗長度解析～		
要 旨			
<p>パラメトリックな離散無記憶情報源に対するユニバーサル符号としては、事前分布を用いて構成するベイズ符号を用いる方法があり、パラメータ集合が絶対連続あるいは離散の場合の漸近的冗長度性能は Clarke&amp;Barron によって精密な解析が与えられている。</p> <p>しかしながら、パラメータ集合が特異なクラスの場合については我が国の研究者によるものを除くと研究がない。</p> <p>最初に問題提起をした Takahashi に続き、より具体的な自己相似測度に従う場合に対して、橋本が上界を、久富が下界を与え結果として一次冗長度性能が（レニー情報次元）<math>\times 1/2 \log n</math> となることを示した。一方、松上は情報源データを生成するパラメータが動的なシステムの出力過程に従うと考えパラメータが未知の事前分布に従ってランダムに生成されるような無記憶情報源モデルを提案した。</p> <p>本稿では松上の概略的な考え方を精緻化することを目指した。具体的にはまず符号構成法を簡素化した。さらに粗く設定されていたエルゴート性に関する仮定を緩和しその下でその符号が過去のサンプル過程の観測数に応じて 1 次冗長度性能を改善することを、特に特異事前分布の場合に示した。</p>			

平成 30 年度 修士学位論文

事前分布の推定に基づく無歪情報源符号  
～符号化アルゴリズムと冗長度解析～

電気通信大学 大学院 情報理工学研究科  
博士前期課程 情報・ネットワーク工学専攻

1731015 五十嵐 棕介

指導教員 川端 勉 教授 大濱 靖匡 教授

提出 平成 31 年 1 月 28 日



## 目次

<b>1</b>	<b>はじめに</b>	<b>2</b>
<b>2</b>	<b>マルコフフラクタル集合</b>	<b>3</b>
2.1	マルコフフラクタル集合 . . . . .	3
2.2	マルコフフラクタル測度 . . . . .	4
2.3	典型集合 . . . . .	4
2.4	Rényi 情報次元 . . . . .	5
<b>3</b>	<b>既知の事前分布に基づくベイズ符号とその冗長度解析</b>	<b>7</b>
3.1	冗長度 . . . . .	7
3.2	Clarke&Barron 理論からの準備 . . . . .	7
3.3	既存の結果 . . . . .	11
<b>4</b>	<b>事前分布の推定に基づくベイズ符号とその冗長度解析</b>	<b>12</b>
4.1	問題設定 . . . . .	12
4.2	主定理 . . . . .	13
4.3	冗長度解析 . . . . .	16
<b>5</b>	<b>まとめ</b>	<b>19</b>

# 第 1 章

## はじめに

パラメトリックな離散無記憶情報源に対するユニバーサル符号としては，事前分布を用いて構成するベイズ符号を用いる方法が有名であり，パラメータ集合が絶対連続あるいは離散の場合の漸近的冗長度性能は Clarke&Barron[1] によって精密な解析が与えられている．しかしながら，文献ではパラメータ集合が特異なクラスの場合については考察されていない．

最初に特異事前分布を用いた場合の無歪ユニバーサル符号化問題を考察したのは Takahashi[2] であり，橋本 [4] はパラメータが特異の中でも自己相似測度に従う場合を考察し，その場合には (Rényi 情報次元)  $\times 1/2 \log n$  という形で上界できることを示唆し，久富 [5] が橋本の上界をより明確化した．さらに下界も同時に与え，両者がタイトであることを示した．

松上 [8] は情報源データを生成するパラメータは動的なシステムの出力過程に従って変動するモデルを考えた．非線形システム論によるとシステムの出力値の集合は離散でも絶対連続でもない場合がある．この場合出力分布は仮定できるとしても未知とする場合が有用であろうと考え，未知の事前分布をそれに基づいて生成されたデータ列から推定し，それに基づいて符号化する方法を考察した．その結果，いくつかの仮定の下でその符号が過去のサンプル過程の観測数に応じて 1 次冗長度性能を改善することを，特に特異事前分布の場合に示した．

本稿では，松上と同様に，未知の事前分布に基づいて生成されたデータから推定した事前分布を用いた符号化を考察するが，符号構成法を改善し，その上でいくつかの仮定を整理し，冗長度解析をより簡潔かつ精緻なものにする．

本稿の構成は以下の通りである．第 2 章では，特異事前分布のモデルとしてのマルコフフラクタル測度について述べ，特にその上の Rényi 情報次元を定義する．第 3 章では，既知のマルコフフラクタル事前分布に基づくベイズ符号の漸近的冗長度の既存研究について述べる．第 4 章では，推定した事前分布に基づくベイズ符号を提案し，マルコフフラクタル事前分布である場合の冗長度解析を行う．

## 第 2 章

# マルコフフラクタル集合

この章では、特異集合の 1 つとしてマルコフフラクタル集合を定義する．そして、その測度としてマルコフフラクタル測度を定義するとともにフラクタル次元として Rényi 情報次元を定義する．

### 2.1 マルコフフラクタル集合

任意の有界集合を  $S \subseteq \mathbb{R}^d$  ( $S \neq \emptyset$ )，いくつかの状態をとり得る状態集合を  $M$  とする．初期状態  $i \in M$  に付随する相似変換を  $W_i$ ，その縮小率を  $\alpha_i^{-1}$  とする．それ以降はマルコフ過程を持ち，1 つ前の状態  $i \in M$  から状態  $j \in M$  への遷移に付随する相似変換を  $W_{ij}$ ，その縮小率を  $\alpha_{ij}^{-1}$  とする．

次に，集合列  $\{S_k\}$  を以下のように定義する．

$$S_1 = \bigcup_{b_1} W_{b_1}(S) \quad (2.1)$$

$$S_2 = \bigcup_{b_1, b_2} W_{b_1} W_{b_1 b_2}(S) \quad (2.2)$$

$\vdots$

$$S_k = \bigcup_{b_1, b_2, \dots, b_k} W_{b_1} W_{b_1 b_2} \cdots W_{b_{k-2} b_{k-1}} W_{b_{k-1} b_k}(S) \quad (2.3)$$

$k \rightarrow \infty$  とした極限集合  $S_\infty$  をマルコフフラクタル集合と呼ぶ．この集合  $S_\infty$  は，状態集合  $M = \{1, \dots, m\}$  上の有限ならびに無限系列の集合  $M^*$  に写像できる．また， $\forall \theta \in S_\infty$  に対して系列  $b(\theta) = b_1 b_2 \cdots b_\infty \in M^*$  が対応する．したがって，集合  $S_\infty$  での確率測度を  $\mu$  とすれば，写像  $\theta \rightarrow b(\theta)$  によって状態集合  $M^*$  上の確率測度  $\mu^*$  が誘導される．

上記の定義のもとで以下に状態数が  $m = 2$  個である場合のマルコフフラクタル集合を示す。

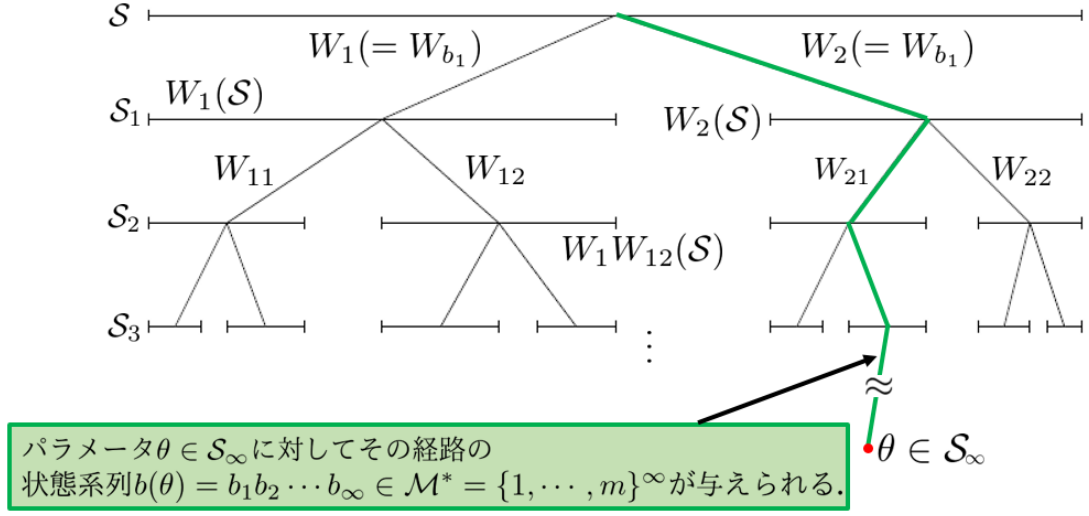


図 2.1: マルコフフラクタル集合 ( $m = 2$  の場合)

## 2.2 マルコフフラクタル測度

任意の集合  $A \subset \mathbb{R}^d$  に対して、初期状態での確率測度を  $\mu(A)$ 、それ以降の状態  $b_1$  での確率測度を  $\mu^{(b_1)}(A)$  とするとき、これらは以下のような再帰関係式を満たす。

$$\mu(A) = \sum_{b_1} p_{b_1} \mu^{(b_1)}(W_{b_1}^{-1}A) \quad (2.4)$$

$$\mu^{(b_1)}(A) = \sum_{b_2} p_{b_1 b_2} \mu^{(b_2)}(W_{b_1 b_2}^{-1}A) \quad (2.5)$$

上記の確率測度  $\mu$  をマルコフフラクタル測度と呼ぶ。

## 2.3 典型集合

マルコフフラクタル測度  $\mu$  に対応する集合  $\mathcal{M}^*$  上の確率測度  $\mu^*$  がエントロピー  $H_\infty(\mu^*)$  をもつ定常エルゴード過程であるとする。  $\mu^*$  の典型集合  $A_\epsilon^k$  を以下で定義する。

$$A_\epsilon^k = \left\{ b_1^k : \left| -\frac{1}{k} \log \mu^*(b_1^k) - H_\infty(\mu^*) \right| < \epsilon \right\} \quad (2.6)$$

また、  $b_1^k = b_1 b_2 \cdots b_k \in A_\epsilon^k$  を典型系列と呼ぶ。

## 2.4 Rényi 情報次元

以降では、マルコフフラクタル集合  $\mathcal{S}_\infty$  を単に集合  $\mathcal{S}$  と記述する。パラメータ  $\theta$  を中心とする半径  $\delta$  のユークリッド球  $B(\theta, \delta)$  を考え、 $B(\theta, \delta) = \{y \in \mathbb{R}^d : \|y - \theta\| \leq \delta\}$  と定義する。次に  $\mathcal{S}$  上の確率分布に対して確率  $\epsilon$  ほどを除いた確率被覆に関する次元  $\dim^{(\epsilon)} \mathcal{S}$  を定義する。これは被覆できない部分の確率が  $\epsilon$  以下となる最小数  $N_{\min}$  個のユークリッド球  $B(\theta, \delta)$  に対する情報レートとして以下のように定義される。

$\forall \epsilon > 0$  に対して、

$$\dim^{(\epsilon)} \mathcal{S} = \lim_{\delta \downarrow 0} \frac{\log N_{\min}}{\log \delta^{-1}} \quad (2.7)$$

$$N_{\min} = \{N : \exists \theta_i \in \mathcal{S} \subset \mathbb{R}^d, \mu(\cup_{i=1}^N B(\theta_i, \delta)) \geq 1 - \epsilon\} \quad (2.8)$$

ここで、典型集合  $A_\epsilon^{k(\delta)}$  に対して、 $N_{\min} \leq \#A_\epsilon^{k(\delta)}$  と書ける。また、確率測度  $\mu^*$  に関する期待値を  $E^*$  と書き、

$$k(\delta) = \left\lfloor \frac{\log \delta^{-1}}{(1 - \epsilon)E^* \log \alpha_{b_1 b_2}^{-1}} \right\rfloor \quad (2.9)$$

と定義すると、以下の上界の式が成り立つ。

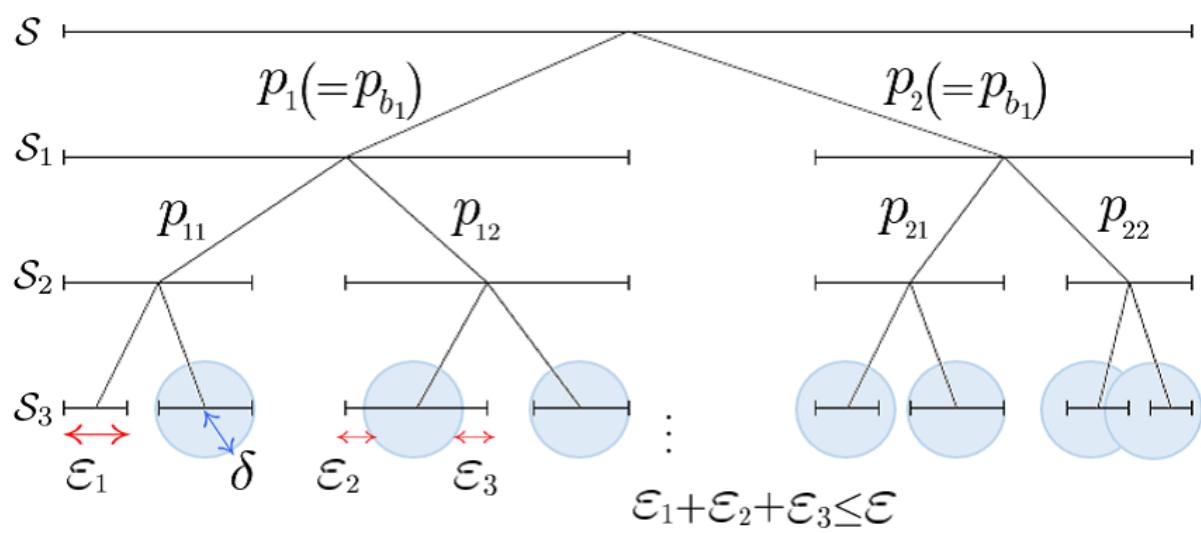
$$\lim_{\epsilon \downarrow 0} \dim^{(\epsilon)} \mathcal{S} \leq \lim_{\epsilon \downarrow 0} \lim_{\delta \downarrow 0} \frac{\log \#A_\epsilon^{k(\delta)}}{\log \delta^{-1}} \quad (2.10)$$

$$\leq \lim_{\epsilon \downarrow 0} \lim_{\delta \downarrow 0} \frac{k(\delta)\{H_\infty(\mu^*) + \epsilon\}}{\log \delta^{-1}} \quad (2.11)$$

$$< \lim_{\epsilon \downarrow 0} \frac{H_\infty(\mu^*) + \epsilon}{(1 - \epsilon)E^* \log \alpha_{b_1 b_2}^{-1}} \quad (2.12)$$

$$= \frac{H_\infty(\mu^*)}{E^* \log \alpha_{b_1 b_2}^{-1}} \quad (2.13)$$

式 (2.13) がマルコフフラクタル測度での Rényi 情報次元  $\rho(\mu)$  となる。下界の評価も上界と同様にできる。

図 2.2: マルコフフラクタル測度の  $(1-\varepsilon)$  被覆



## 第 3 章

# 既知の事前分布に基づくベイズ符号とその冗長度解析

この章では、既知の事前分布に基づくベイズ符号とその冗長度解析の既存研究について述べる。その中でも、文献 [1] における議論が基本となっているため、それに沿って必要な仮定や条件を述べた上で、本研究でも用いる補題や既存研究の結果について述べる。

### 3.1 冗長度

アルファベット集合を  $\mathcal{X}$ ，パラメータ集合を  $\Theta \subset \mathbb{R}^d$  とし，パラメータ  $\theta \in \Theta$  における長さ  $n$  のデータ  $x^n \in \mathcal{X}^n$  の確率分布を  $p(x^n|\theta)$  とする。パラメータの事前分布を  $w(\theta)$  とすると，ベイズ混合分布  $m(x^n)$  は  $m(x^n) = \int_{\Theta} p(x^n|\theta)w(\theta)d\theta$  と表される。この負対数をとった  $-\log m(x^n)$  がベイズ符号の符号長であり，この符号長から真のパラメータ  $\theta_0$  における符号長  $-\log p(x^n|\theta_0)$  を差し引いたものを冗長度  $R$  と定義する。

$$R = -\log m(x^n) - \{-\log p(x^n|\theta_0)\} \quad (3.1)$$

### 3.2 Clarke&Barron 理論からの準備

文献 [1] を参考にして，以下に正則条件を定義していく。

条件 1.

確率分布  $p(x|\theta)$  がほとんど全てのデータ  $x$  に対して真のパラメータ  $\theta_0$  で 2 階微分可能である。また 1 から  $d$  まで取りうる変数  $j, k$  において  $\delta > 0$  が存在し，次式が成り立つ。

$$E \sup_{\|\theta - \theta_0\| < \delta} \left| \frac{\partial^2}{\partial \theta_j \partial \theta_k} \log p(X|\theta) \right|^2 < \infty \quad (3.2)$$

$$E \left| \frac{\partial}{\partial \theta_j} \log p(X|\theta_0) \right|^2 < \infty \quad (3.3)$$

**条件 2.**

$\theta_0$  におけるダイバージェンス  $D(p(x|\theta)||p(x|\theta_0))$  について  $\theta$  に関する 2 次導関数

$$J_{\theta_0} = J(\theta_0) = \left[ \frac{\partial^2}{\partial \theta_j \partial \theta_k} D(p(x|\theta)||p(x|\theta_0)) \right]_{j,k=1 \dots d} \quad (3.4)$$

が存在する.

**条件 3.**

$w(\cdot|X^n)$  を  $X^n$  が与えられたもとでの  $\theta$  の事後確率とすると、次式が成り立つ.

$$P^n\{w(N_\delta^c|X^n) > \delta\} = o(1/\log n) \quad (3.5)$$

その他, 経験情報行列を

$$J_\theta^* = J^*(\theta) = - \left[ \frac{1}{n} \frac{\partial^2}{\partial \theta_j \partial \theta_k} (\log p(X^n|\theta)) \right]_{j,k=1 \dots d} \quad (3.6)$$

平均スコア関数, 規格化スコア関数を

$$l'_n(\theta_0) = \frac{1}{n} \nabla \log p(X^n|\theta_0), \quad S_n(\theta_0) = \sqrt{n} \cdot l'_n \quad (3.7)$$

と定義する. すると, 最尤推定量  $\hat{\theta}$  はほぼ以下の式で表現できる.

$$\hat{\theta} = \theta_0 + J_{\theta_0}^{-1} l'_n(\theta_0) \quad (3.8)$$

真のパラメータ  $\theta_0$  を中心とし, 半径  $\delta$  のフィッシャー情報球  $N_\delta$  を次式のような  $\theta$  の集まりと定義する.

$$N_{\delta(n)} = N(\theta_0, \delta(n)) = \{\theta : |\theta - \theta_0| \leq \delta(n)\} \quad (3.9)$$

ただし,  $|\theta - \theta_0|^2 = |\theta - \theta_0|_{J_{\theta_0}}^2 = |\theta - \theta_0|^T J_{\theta_0} |\theta - \theta_0|$

$$\delta(n) = n^{-\frac{1}{2} + o_P(1)}$$

さらに, 任意の  $0 < \varepsilon < 1, \delta > 0$  において, 次の典型的な事象  $A_n, B_n, C_n$  を定義する.

$$A_n(\delta, \varepsilon) = \left\{ \int_{N_\delta^c} p(X^n|\theta)w(\theta)d\theta \leq \varepsilon \int_{N_\delta} p(X^n|\theta)w(\theta)d\theta \right\} \quad (3.10)$$

$$\begin{aligned} B_n(\delta, \varepsilon) &= \left\{ (1 - \varepsilon)(\theta - \theta_0)^T J_{\theta_0}(\theta - \theta_0) \right. \\ &\leq (\theta - \theta_0)^T J_{\tilde{\theta}}^*(\theta - \theta_0) \\ &\leq (1 + \varepsilon)(\theta - \theta_0)^T J_{\theta_0}(\theta - \theta_0), \\ &\quad \left. \text{for all } \theta, \tilde{\theta} \in N_\delta \right\} \end{aligned} \quad (3.11)$$

$$C_n(\delta) = \{l'_n(\theta_0)^T J_{\theta_0}^{-1} l'_n(\theta_0) \leq \delta^2\} \quad (3.12)$$

$A_n$  は近傍  $N_\delta$  の事後確率が少なくとも  $1/(1 + \varepsilon)$  であるようなデータ  $x^n$  の集合である.

$B_n$  は経験情報行列がダイバージェンスの2次導関数行列を用いて上下界される事象である.

$C_n$  は平均スコアのノルムがほぼ0である事象である.

仮定 1. ( $S$  の分離条件)

$\theta$  を中心とする半径  $u$  の球の内部を  $B(\theta, u)$  とする. 各  $\theta_0 \in \Theta$  に対して  $M = \{1, \dots, m\}$  に属する数列  $\{b_k\}_{k=1}^\infty$  を  $(\forall k)(\theta_0 \in W_{b_1} \dots W_{b_k} S)$  となるように定め, 相似変換により常に  $\theta_0$  の方向へ遷移する場合, 以下の  $\Delta k$  が有限であると仮定する.

$$\Delta k = \sup_u \left\{ \sup_{\theta_0 \in S} \left\{ \inf \{k: B(\theta, u) \cap S \supseteq W_{b_1} \dots W_{b_k} S\} - \sup \{k: B(\theta, u) \cap S \subseteq W_{b_1} \dots W_{b_k} S\} \right\} \right\} \quad (3.13)$$

右辺の第一項は  $S$  の縮小領域が  $\theta_0$  の近傍を含む最大の縮小回数を表し, 第二項は  $S$  の縮小領域が  $\theta_0$  の近傍に含まれる最小の縮小回数を表す.  $\Delta k$  はその両者の差の最大値を表している.

補題 1.

正則条件の下で, 任意の  $0 < \varepsilon < 1, \delta > 0$  に対して,

$$\frac{m(x^n)}{p(x^n|\theta_0)} \leq (1 + \varepsilon) \cdot e^{\{\frac{1}{2(1-\varepsilon)} S_n^T J_{\theta_0}^{-1} S_n\}} \cdot \int_{N_{\delta(n)}} w(d\theta) \quad (3.14)$$

また,

$$\frac{m(x^n)}{p(x^n|\theta_0)} \geq e^{\{\frac{1}{2(1+\varepsilon)} S_n^T J_{\theta_0}^{-1} S_n\}} \cdot e^{-\frac{n\varepsilon^2}{2(1+\varepsilon)} \delta(n)^2} \cdot \int_{N_{\delta(n)}} w(\theta) d\theta \quad (3.15)$$

が成り立つ.

補題 2.

正則条件の下で, 任意の連続事前分布  $w(\theta)$  に対して次式が成り立つ.

$$-\log \int_{N_\delta} w(d\theta) \sim \frac{d}{2} \log n + o_P(\log n) \quad (3.16)$$

補題 3.

正則条件と仮定 1 の下で, 特異事前分布  $w_s(\theta)$  に対して次式が成り立つ.

$$-\log \int_{N_\delta} w_s(d\theta) \sim \frac{\rho(\mu)}{2} \log n + o_P(\log n) \quad (3.17)$$

ここでは, 以下の stochastic オーダーを用いた.

$$Y_n = o_P(f(n)), as n \rightarrow \infty \Leftrightarrow (\forall \epsilon) \left( \lim_{n \rightarrow \infty} P(|Y_n| < \epsilon |f(n)|) = 1 \right) \quad (3.18)$$

### 3.3 既存の結果

定理 1. [1]

正則条件の下で次式が成り立つ.

$$\lim_{n \rightarrow \infty} \left( -\log \frac{m(x^n)}{p(x^n|\theta_0)} - \frac{d}{2} \log \frac{n}{2\pi} + \frac{1}{2} S_n^T J_{\theta_0}^{-1} S_n \right) = \log \frac{1}{w(\theta)} + \frac{1}{2} \log \det J_{\theta_0} \quad (3.19)$$

定理 2. [6]

正則条件と仮定 1 の下で, 事前分布  $w(\theta)$  がマルコフフラクタル事前分布であるとする  
と次式が成り立つ.

$$\lim_{n \rightarrow \infty} \left\{ \frac{1}{\log n} \left( -\log \frac{m(x^n)}{p(x^n|\theta_0)} - \frac{\rho(\mu)}{2} \right) \right\} = 0 \quad (3.20)$$

## 第 4 章

# 事前分布の推定に基づくベイズ符号とその冗長度解析

この章では、既存研究とは異なり、パラメータが事前分布に従い非定常に変化する場合を考えていく。はじめに問題設定を述べ、その後いくつかの仮定をおいた上での冗長度解析を行う。

### 4.1 問題設定

確率モデルを指定するパラメータ集合  $\Theta \subset \mathbb{R}^d$  上にある力学系  $\theta_{i+1} = f(\theta_i)$  に従って、未知のパラメータ列  $\theta_i \in \Theta$  ( $i = 1, 2, \dots, k$ ) が生成され、この力学系は可測な集合  $\mathcal{A} \subset \Theta$  に対してパラメータの観測数  $k$  が十分大きいとき以下のエルゴード性が満たされると仮定する。

$$\lim_{k \rightarrow \infty} \frac{1}{k} \sum_{i=1}^k 1\{\theta_i \in \mathcal{A}\} \rightarrow w(\mathcal{A}) \quad (4.1)$$

未知のパラメータ  $\theta_i$  それぞれに対して長さ  $l$  のデータ  $X_{(i)}^l \sim p(x^l|\theta_i) = \prod_{j=1}^l p(x_j|\theta_i)$  を観測したとき、観測したデータから最尤推定量  $\hat{\theta}_i(X_{(i)}^l)$  が得られるもととする。、無情報事前分布として Jeffreys 事前分布  $w_J(\theta)$  を用いた場合のベイズ混合分布を  $m_s(x^n) = \int_{\Theta} p(x^n|\theta)w_s(\theta)d\theta$  と表記し、特異事前分布  $w_s(\theta)$  を用いた場合も同様に表記することにする。

そして、データ列  $x_{(1)}^l, x_{(2)}^l, \dots, x_{(k)}^l$  に基づく、新たなデータ列  $x^n$  の確率推定量  $\hat{m}(x^n)$  を以下のように定義する。

$$\hat{m}(x^n) = (1 - \alpha(k)) \frac{1}{k} \sum_{i=1}^k m_J(x^n|x_{(i)}^l) + \alpha(k)m_J(x^n) \quad (4.2)$$

このとき、 $0 \leq \alpha(k) \leq 1$  はデータ列数  $k$  の単調減少関数であり、詳しくは後に定義する。

この確率推定量を用いると、評価したい冗長度  $R$  は以下のような関数となる.

$$R = -\log \left\{ (1 - \alpha(k)) \frac{1}{k} \sum_{i=1}^k \frac{m_J(x^n | x_{(i)}^l)}{p(x^n | \theta_0)} + \alpha(k) \frac{m_J(x^n)}{p(x^n | \theta_0)} \right\} \quad (4.3)$$

## 4.2 主定理

冗長度の第1項の極限は以下のように上下界できる.

**定理 3.**

正則条件とエルゴード性のもとで、データ長  $l, n$  に関して漸近的に以下の不等式が成立する.

$$\begin{aligned} & \lim_{k \rightarrow \infty} \frac{1}{k} \sum_{i=1}^k \frac{m_J(x^n | x_{(i)}^l)}{p(x^n | \theta_0)} \\ & \leq (1 + \epsilon)^2 e^{\frac{\epsilon^2}{2(1+\epsilon)}} \left( \frac{l+n}{l} \right)^{\frac{d}{2} + o_P(1)} \left\{ \frac{2(1 - 2\epsilon - \epsilon^2)(l+n) - (1 + \epsilon)l}{4\pi(1 - \epsilon^2)(l+n)} \right\}^{-\frac{d}{2}} \\ & \quad \cdot \exp \left\{ \frac{(1 - 2\epsilon - \epsilon^2)n}{2(1 - \epsilon)\{(1 - 2\epsilon - \epsilon^2)(l+n) - (1 + \epsilon)l\}} S_n^T J_{\theta_0}^{-1} S_n \right\} \cdot \int_{N_{\delta(l)}(\theta_0)} w_s(d\theta) \end{aligned} \quad (4.4)$$

同様に、以下の不等式も成立する.

$$\begin{aligned} & \lim_{k \rightarrow \infty} \frac{1}{k} \sum_{i=1}^k \frac{m_J(x^n | x_{(i)}^l)}{p(x^n | \theta_0)} \\ & \geq \frac{e^{\frac{\epsilon^2}{2(1+\epsilon)}}}{(1 + \epsilon)} \left( \frac{l+n}{l} \right)^{\frac{d}{2} + o_P(1)} \left\{ \frac{(1 + 2\epsilon - \epsilon^2)(l+n) - (1 - \epsilon)l}{(1 - \epsilon^2)(l+n)} \right\}^{-\frac{d}{2}} \\ & \quad \cdot \exp \left\{ \frac{(1 + 2\epsilon - \epsilon^2)n}{2(1 + \epsilon)\{(1 + 2\epsilon - \epsilon^2)(l+n) - (1 - \epsilon)l\}} S_n^T J_{\theta_0}^{-1} S_n \right\} \cdot \int_{N_{\delta(l)}(\theta_0)} w_s(d\theta) \end{aligned} \quad (4.5)$$

**証明.**

$$\lim_{k \rightarrow \infty} \frac{1}{k} \sum_{i=1}^k \frac{m_J(x^n | x_{(i)}^l)}{p(x^n | \theta_0)} = \lim_{k \rightarrow \infty} \frac{1}{k} \sum_{i=1}^k \sum_{x^l} \frac{m_J(x^n | X^l) p(X^l | \theta_i)}{p(x^n | \theta_0)} \quad (4.6)$$

$$\rightarrow \sum_{x^l} \int_{\Theta} \frac{m_J(x^n | x^l)}{p(x^n | \theta_0)} p(x^l | \theta) w_s(\theta) d\theta \quad (4.7)$$

$$= \sum_{x^l} \frac{m_J(x^n | x^l) m_s(x^l)}{p(x^n | \theta_0)} \quad (4.8)$$

$$= \sum_{x^l} p(x^l | \theta_0, x^n) \cdot \frac{\frac{m_s(x^l)}{p(x^l | \theta_0)}}{\frac{m_J(x^l)}{p(x^l | \theta_0)}} \cdot \frac{m_J(x^l x^n)}{p(x^l x^n | \theta_0)} \quad (4.9)$$

これを補題1, 2を用いて上界すると,

$$\begin{aligned} & \sum_{x^l} p(x^l | \theta_0, x^n) \cdot \frac{\frac{m_s(x^l)}{p(x^l | \theta_0)}}{\frac{m_J(x^l)}{p(x^l | \theta_0)}} \cdot \frac{m_J(x^l x^n)}{p(x^l x^n | \theta_0)} \\ & \leq \sum_{x^l} p(x^l | \theta_0, x^n) \cdot \frac{(1 + \epsilon) e^{\frac{1}{2(1-\epsilon)} S_l^T J_{\theta_0}^{-1} S_l} \int_{N_{\delta(l)}(\theta_0)} w_s(d\theta)}{e^{-\frac{\epsilon^2}{2(1+\epsilon)} \delta(l)^2} e^{\frac{1}{2(1+\epsilon)} S_l^T J_{\theta_0}^{-1} S_l} \int_{N_{\delta(l)}(\theta_0)} w_J(d\theta)} \cdot (1 + \epsilon) e^{\frac{1}{2(1-\epsilon)} S_{l+n}^T J_{\theta_0}^{-1} S_{l+n}} \int_{N_{\delta(l+n)}(\theta_0)} w_J(d\theta) \end{aligned} \quad (4.10)$$

$$= \sum_{x^l} p(x^l | \theta_0) \cdot e^{\frac{\epsilon}{1-\epsilon^2} S_l^T J_{\theta_0}^{-1} S_l} \cdot e^{\frac{1}{2(1-\epsilon)} S_{l+n}^T J_{\theta_0}^{-1} S_{l+n}} \cdot (1 + \epsilon)^2 e^{\frac{\epsilon^2}{2(1+\epsilon)} \left( \frac{l+n}{l} \right)^{\frac{d}{2} + o_P(1)}} \int_{N_{\delta(l)}(\theta_0)} w_s(d\theta) \quad (4.11)$$

となる. ここで,

$$\begin{aligned} S_{l+n}^T J_{\theta_0}^{-1} S_{l+n} & \sim \left( \frac{1}{\sqrt{l+n}} \nabla \log p(x^l x^n | \theta_0) \right)^T J_{\theta_0}^{-1} \left( \frac{1}{\sqrt{l+n}} \nabla \log p(x^l x^n | \theta_0) \right) \\ & = \left\{ \frac{1}{\sqrt{l+n}} \nabla (\log p(x^l | \theta_0) + \log p(x^n | \theta_0)) \right\}^T J_{\theta_0}^{-1} \left\{ \frac{1}{\sqrt{l+n}} \nabla (\log p(x^l | \theta_0) + \log p(x^n | \theta_0)) \right\} \\ & = \left( \sqrt{\frac{l}{l+n}} S_l + \sqrt{\frac{n}{l+n}} S_n \right)^T J_{\theta_0}^{-1} \left( \sqrt{\frac{l}{l+n}} S_l + \sqrt{\frac{n}{l+n}} S_n \right) \end{aligned}$$

であることより, 簡単のために,

$$\begin{aligned} Z &:= J_{\theta_0}^{-\frac{1}{2}} S_l \\ W &:= J_{\theta_0}^{-\frac{1}{2}} S_n \end{aligned}$$

と置くと, 式(4.11)は,

$$\begin{aligned} & \sum_{x^l} p(x^l | \theta_0) \cdot e^{\frac{\epsilon}{1-\epsilon^2} S_l^T J_{\theta_0}^{-1} S_l} \cdot e^{\frac{1}{2(1-\epsilon)} S_{l+n}^T J_{\theta_0}^{-1} S_{l+n}} \cdot (1 + \epsilon)^2 e^{\frac{\epsilon^2}{2(1+\epsilon)} \left( \frac{l+n}{l} \right)^{\frac{d}{2} + o_P(1)}} \int_{N_{\delta(l)}(\theta_0)} w_s(d\theta) \\ & = \int \frac{1}{\sqrt{2\pi^d}} e^{-\frac{1}{2} Z^T Z} \cdot e^{\frac{\epsilon}{1-\epsilon^2} Z^T Z} \cdot e^{\frac{1}{2(1-\epsilon)} \left( \frac{l}{l+n} Z + \frac{n}{l+n} W \right)^T \left( \frac{l}{l+n} Z + \frac{n}{l+n} W \right)} dZ \cdot (1 + \epsilon)^2 \left( \frac{l+n}{l} \right)^{\frac{d}{2} + o_P(1)} \int_{N_{\delta(l)}(\theta_0)} w_s(d\theta) \end{aligned} \quad (4.12)$$

と書ける. ここで更に以下のように置くとすると,

$$\begin{aligned} \alpha &= \frac{\epsilon}{1 - \epsilon^2} \\ \beta &= \frac{1}{2(1 - \epsilon)} \\ \gamma &= \sqrt{\frac{l}{l+n}} \\ \bar{\gamma} &= \sqrt{\frac{n}{l+n}} \end{aligned} \quad (4.13)$$



式(4.12)の積分部は、ガウス積分を用いることにより、

$$\begin{aligned} & \int \exp \left\{ -\frac{1}{2} Z^T Z + \frac{\epsilon}{1-\epsilon^2} Z^T Z + \frac{1}{2(1-\epsilon)} \left( \sqrt{\frac{l}{l+n}} Z + \sqrt{\frac{n}{l+n}} W \right)^T \left( \sqrt{\frac{l}{l+n}} Z + \sqrt{\frac{n}{l+n}} W \right) \right\} dZ \\ &= \int \exp \left\{ -\frac{1}{2} (1 - 2(\alpha + \beta\gamma^2)) Z^T Z + (2\beta\gamma\hat{\gamma}) W^T Z + \beta\hat{\gamma}^2 W^T W \right\} dZ \end{aligned} \quad (4.14)$$

$$= \sqrt{2\pi}^d \cdot \det |(1 - 2(\alpha + \beta\gamma^2))I|^{-\frac{1}{2}} \cdot \exp \left\{ \left( \frac{(2\beta\gamma\hat{\gamma})^2}{2} \cdot \frac{1}{1 - 2(\alpha + \beta\gamma^2)} + \beta\hat{\gamma}^2 \right) W^T W \right\} \quad (4.15)$$

となる。この指数部を計算すると、

$$\frac{(2\beta\gamma\hat{\gamma})^2}{2} \cdot \frac{1}{1 - 2(\alpha + \beta\gamma^2)} + \beta\hat{\gamma}^2 = \frac{(1 - 2\alpha)\beta\hat{\gamma}^2}{1 - 2\alpha - 2\beta\gamma^2} \quad (4.16)$$

$$= \frac{\left(1 - \frac{2\epsilon}{1-\epsilon^2}\right) \frac{n}{2(1-\epsilon)(l+n)}}{\left(1 - \frac{2\epsilon}{1-\epsilon^2}\right) - \frac{l^2}{2(1-\epsilon)(l+n)}} \quad (4.17)$$

$$= \frac{(1 - 2\epsilon - \epsilon^2) n}{2(1 - \epsilon)\{(1 - 2\epsilon - \epsilon^2)(l + n) - (1 + \epsilon) l\}} \quad (4.18)$$

となることより、式(4.15)は、

$$\begin{aligned} & \sqrt{2\pi}^d \cdot \det |(1 - 2(\alpha + \beta\gamma^2))I|^{-\frac{1}{2}} \cdot \exp \left\{ \left( \frac{(2\beta\gamma\hat{\gamma})^2}{2} \cdot \frac{1}{1 - 2(\alpha + \beta\gamma^2)} + \beta\hat{\gamma}^2 \right) W^T W \right\} \\ &= \left\{ \frac{2(1 - 2\epsilon - \epsilon^2)(l + n) - (1 + \epsilon) l}{4\pi(1 - \epsilon^2)(l + n)} \right\}^{-\frac{d}{2}} \cdot \exp \left\{ \frac{(1 - 2\epsilon - \epsilon^2) n}{2(1 - \epsilon)\{(1 - 2\epsilon - \epsilon^2)(l + n) - (1 + \epsilon) l\}} W^T W \right\} \end{aligned} \quad (4.19)$$

となる。したがって式(4.11)は、

$$\begin{aligned} & \int \frac{1}{\sqrt{2\pi}^d} e^{-\frac{1}{2} Z^T Z} \cdot e^{\frac{\epsilon}{1-\epsilon^2} Z^T Z} \cdot e^{\frac{1}{2(1-\epsilon)} \left( \frac{l}{l+n} Z + \frac{n}{l+n} W \right)^T \left( \frac{l}{l+n} Z + \frac{n}{l+n} W \right)} dZ \cdot (1 + \epsilon)^2 \left( \frac{l + n}{l} \right)^{\frac{d}{2} + o(1)} \int_{N_\delta(\theta_0)} w_s(d\theta) \\ &= (1 + \epsilon)^2 e^{\frac{\epsilon^2}{2(1+\epsilon)}} \left( \frac{l + n}{l} \right)^{\frac{d}{2} + o_P(1)} \left\{ \frac{2(1 - 2\epsilon - \epsilon^2)(l + n) - (1 + \epsilon) l}{4\pi(1 - \epsilon^2)(l + n)} \right\}^{-\frac{d}{2}} \\ & \quad \cdot \exp \left\{ \frac{(1 - 2\epsilon - \epsilon^2) n}{2(1 - \epsilon)\{(1 - 2\epsilon - \epsilon^2)(l + n) - (1 + \epsilon) l\}} S_n^T J_{\theta_0}^{-1} S_n \right\} \cdot \int_{N_\delta(l)(\theta_0)} w_s(d\theta) \end{aligned} \quad (4.20)$$

となる。

下界の証明も同様の計算によって得られる。

□

### 4.3 冗長度解析

以降は  $l = n$  として議論を行っていく．初めに  $\alpha(k)$  の定義を行う．

**定義 1.** ( $\alpha(k)$  の定義)

$\alpha(k)$  を以下のように定義する．

$$\alpha(k) = e^{-\frac{k}{k_0}} \quad (4.21)$$

ただし,  $k_0 \in \mathbb{N}$  は以下の条件を満たすように定める．

$$\Pr \left( \exists k' \geq k_0, \left| \frac{\frac{1}{k'} \sum_{i=1}^{k'} \frac{m_J(x^n | X_{(i)}^d)}{p(x^n | \theta_0)}}{E_{p(x^l | \theta)} \frac{m_J(x^n | X^l)}{p(x^n | \theta_0)}} - 1 \right| > \frac{1}{2} \right) = o_P(1) \quad (4.22)$$

および, ある  $\nu > 0$  に対して,

$$k_0 \geq e^{(\frac{d-\rho}{2} + \nu) \log n} \quad (4.23)$$

を満たす．

データ列数  $k$  の値によって場合分けをして冗長度解析を行うと以下のような漸近的表現が得られる。

$k \gg k_0$  の場合

$$R = -\log \left\{ (1 - \alpha(k)) \frac{1}{k} \sum_{i=1}^k \frac{m_J(x^n | x_{(i)}^l)}{p(x^n | \theta_0)} + \alpha(k) \frac{m_J(x^n)}{p(x^n | \theta_0)} \right\} \quad (4.24)$$

$$= -\log (1 - \alpha(k)) \frac{1}{k} \sum_{i=1}^k \frac{m_J(x^n | x_{(i)}^l)}{p(x^n | \theta_0)} - \log \left( 1 + \frac{\alpha(k) \frac{m_J(x^n)}{p(x^n | \theta_0)}}{(1 - \alpha(k)) \frac{1}{k} \sum_{i=1}^k \frac{m_J(x^n | x_{(i)}^l)}{p(x^n | \theta_0)}} \right) \quad (4.25)$$

ここで,

$$\epsilon = \frac{\alpha(k) \frac{m_J(x^n)}{p(x^n | \theta_0)}}{(1 - \alpha(k)) \frac{1}{k} \sum_{i=1}^k \frac{m_J(x^n | x_{(i)}^l)}{p(x^n | \theta_0)}} \quad (4.26)$$

と置くと、テーラーの定理より

$$\log(1 + \epsilon) = \epsilon - \frac{1}{2}(\epsilon\xi)^2, \quad (0 < \xi < 1) \quad (4.27)$$

であるので、式 (4.22) と定理 3 を用いて以下のように計算できる。

$$\begin{aligned} E_{p(x^l | \theta)} |\log(1 + \epsilon) - \epsilon| &\leq \frac{1}{2} E_{p(x^l | \theta)} \epsilon^2 \\ &= \frac{1}{2} \left( \frac{\alpha(k)}{1 - \alpha(k)} \right)^2 \cdot E_{p(x^l | \theta)} \left( \frac{\frac{m_J(x^n)}{p(x^n | \theta_0)}}{\frac{1}{k} \sum_{i=1}^k \frac{m_J(x^n | x_{(i)}^l)}{p(x^n | \theta_0)}} \right)^2 \\ &\leq \frac{1}{2} \left( \frac{\alpha(k)}{1 - \alpha(k)} \right)^2 \cdot 4 \left( \frac{\frac{m_J(x^n)}{p(x^n | \theta_0)}}{E_{p(x^l | \theta)} \frac{m_J(x^n | x_{(i)}^l)}{p(x^n | \theta_0)}} \right)^2 \\ &\sim \frac{1}{2} \left( \frac{\alpha(k)}{1 - \alpha(k)} \right)^2 \cdot 4e^{(\rho-d)\log n} \\ &\rightarrow 0 \end{aligned} \quad (4.28)$$

従って,

$$\begin{aligned} R &= -\log (1 - \alpha(k)) \frac{1}{k} \sum_{i=1}^k \frac{m_J(x^n | x_{(i)}^l)}{p(x^n | \theta_0)} - \log \left( 1 + \frac{\alpha(k) \frac{m_J(x^n)}{p(x^n | \theta_0)}}{(1 - \alpha(k)) \frac{1}{k} \sum_{i=1}^k \frac{m_J(x^n | x_{(i)}^l)}{p(x^n | \theta_0)}} \right) \\ &\sim -\log \frac{1}{k} \sum_{i=1}^k \frac{m_J(x^n | x_{(i)}^l)}{p(x^n | \theta_0)} \end{aligned} \quad (4.29)$$

$$\sim \frac{\rho}{2} \log n + o_P(\log n) \quad (4.30)$$

となる。

$k \ll e^{\nu \log n}$  の場合

$$R = -\log \left\{ (1 - \alpha(k)) \frac{1}{k} \sum_{i=1}^k \frac{m_J(x^n | x_{(i)}^l)}{p(x^n | \theta_0)} + \alpha(k) \frac{m_J(x^n)}{p(x^n | \theta_0)} \right\} \quad (4.31)$$

$$= -\log \alpha(k) \frac{m_J(x^n)}{p(x^n | \theta_0)} - \log \left( 1 + \frac{(1 - \alpha(k)) \frac{1}{k} \sum_{i=1}^k \frac{m_J(x^n | x_{(i)}^l)}{p(x^n | \theta_0)}}{\alpha(k) \frac{m_J(x^n)}{p(x^n | \theta_0)}} \right) \quad (4.32)$$

同様に,

$$\epsilon = \frac{(1 - \alpha(k)) \frac{1}{k} \sum_{i=1}^k \frac{m_J(x^n | x_{(i)}^l)}{p(x^n | \theta_0)}}{\alpha(k) \frac{m_J(x^n)}{p(x^n | \theta_0)}} \quad (4.33)$$

と置くと, 式 (4.23) と定理 3 を用いて,

$$|\log(1 + \epsilon') - \epsilon'| \leq \frac{1}{2} \left( \frac{1 - \alpha}{\alpha} \right)^2 \cdot \frac{1}{4} e^{(d-\rho) \log n} \quad (4.34)$$

$$\rightarrow 0 \quad (4.35)$$

と計算できる. ここでは以下の関係を用いた

$$\left( \frac{1 - \alpha}{\alpha} \right) e^{\frac{d-\rho}{2} \log n} \approx (1 - \alpha) e^{\frac{d-\rho}{2} \log n} \quad (4.36)$$

$$\approx \frac{k}{k_0} e^{\frac{d-\rho}{2} \log n} \quad (4.37)$$

$$\leq k e^{-\nu \log n} \quad (4.38)$$

従って,

$$\begin{aligned} R &= -\log \alpha(k) \frac{m_J(x^n)}{p(x^n | \theta_0)} - \log \left( 1 + \frac{(1 - \alpha(k)) \frac{1}{k} \sum_{i=1}^k \frac{m_J(x^n | x_{(i)}^l)}{p(x^n | \theta_0)}}{\alpha(k) \frac{m_J(x^n)}{p(x^n | \theta_0)}} \right) \\ &\sim -\log \frac{m_J(x^n)}{p(x^n | \theta_0)} \end{aligned} \quad (4.39)$$

$$\sim \frac{d}{2} \log n + o_P(\log n) \quad (4.40)$$

となる.

## 第 5 章

### まとめ

本稿では，パラメータが事前分布に従い非定常に変化するという問題設定上で，新たなベイズ符号を提案し，その符号が過去のデータ列数に応じて 1 次冗長度性能を改善することを，特にマルコフフラクタル事前分布に対して示した．

今後の課題としては，冗長度解析を行う際に事前分布のクラスが分からないといけないことや，データ長に関する条件がない場合についての議論などが挙げられる．

## 発表実績

1. 五十嵐棕介, 川端勉, “事前分布の推定に基づく無歪情報源符号の冗長度解析～エルゴード性に関する仮定の緩和～,” 電子情報通信学会技術研究報告, pp. 13–18, Mar. 2018.
2. 五十嵐棕介, 川端勉, “事前分布の推定に基づく無歪情報源符号～符号化アルゴリズムと冗長度解析～,” Proceedings of the 41th Symposium on Information Theory and its Applications, pp. 413–417, Dec. 2018.

# 謝辞

最後に，本研究を行うにあたりお忙しい中時間を割き，丁寧に指導してくださった川端勉教授に深く感謝致します．また，お忙しい中，ゼミ等において大変お世話になりました大濱靖匡教授，八木秀樹准教授，Santoso Bagus 助教に感謝致します．そして，共に研究に励んできた川端研究室，八木研究室的皆様に感謝いたします．本当にありがとうございました．

## 参考文献

- [1] B. Clarke and A. Barron, “Information-theoretic asymptotics of Bayes methods,” *IEEE Trans. Inform. Theory*, vol. 36, pp. 453–471, May. 1990.
- [2] H. Takahashi, “Redundancy of Universal Coding, Kolmogorov Complexity, and Hausdorff Dimension,” *IEEE Trans. Inform. Theory*, vol. 50, No. 11, pp. 2727–2736, Nov. 2004.
- [3] 知念 泰之, “マルコフフラクタル測度に対するレート歪み次元,” 電気通信大学修士論文, Mar. 2007.
- [4] 橋本 達矢, “特異事前分布をもつユニバーサル情報源符号化の冗長度解析,” 電気通信大学修士論文, Mar. 2010.
- [5] 久富 彰人, 竹内 啓悟, 川端 勉, “自己相似な事前分布による無歪情報源符号化の漸近的冗長度,” *Proc. SITA2011*, pp. 88-91, Nov. 2011.
- [6] 松上 直矢, 川端 勉, “マルコフ・フラクタル事前分布をもつ情報源に対するユニバーサル符号の冗長度解析,” 電子情報通信学会技術研究報告, pp. 13–17, Mar. 2016.
- [7] 松上 直矢, 川端 勉, “事前分布の推定に基づくユニバーサル無歪情報源符号の冗長度解析,” *Proceedings of the 39th Symposium on Information Theory and its Applications*, pp. 13–16, Dec. 2016.
- [8] 松上 直矢, 川端 勉, “事前分布の推定に基づく無歪情報源符号の冗長度解析,” 電気通信大学修士論文, Mar. 2017.
- [9] 坂口 浩章, 川端 勉, “有限窓を用いた文脈木重みづけ法,” 電子情報通信学会論文誌, A Vol. J80-A, No.12, pp. 2155–2163, Nov. 1997.